AUTOMATIC CLASSIFICATION METHOD AND APPARATUS

ABSTRACT

Methods and apparatus for classifying an instance (i.e., a data item or a record) automatically into one or more classes selected from a set of potential classes are described. This multi-class classification approach enables instances to be classified with high accuracy, even when the number of classes is very large. In accordance with one classification scheme, a subset of two or more classes to which the instance is determined to most likely belong is selected from the set of potential classes. A classifier that is generated from a set of training records corresponding to a class set inclusive of the selected subset of classes is applied to the instance a to identify at least one class to which the instance most likely belongs.